

***Macromolecular
Sequences in
Systematic and
Evolutionary Biology***

MONOGRAPHS IN EVOLUTIONARY BIOLOGY

Series Editors:

MAX K. HECHT

*Queens College of the
City University of New York
Flushing, New York*

BRUCE WALLACE

*Virginia Polytechnic Institute
and State University
Blacksburg, Virginia*

GHILLEAN T. PRANCE

*New York Botanical Garden
Bronx, New York*

MACROMOLECULAR SEQUENCES IN SYSTEMATIC AND EVOLUTIONARY BIOLOGY

Edited by Morris Goodman

Macromolecular Sequences in Systematic and Evolutionary Biology

Edited by

MORRIS GOODMAN

*Wayne State University
Detroit, Michigan*

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Main entry under title:

Macromolecular sequences in systematic and evolutionary biology.

(Monographs in evolutionary biology)

Includes bibliographical references and index.

1. Chemical evolution. 2. Biology—Classification. 3. Macromolecules. I. Goodman, Morris. 1925— II. Series.

QH371.M27 1982 575 82-15138

ISBN 978-1-4684-4285-4 ISBN 978-1-4684-4283-0 (cBook)

DOI 10.1007/978-1-4684-4283-0

©1982 Plenum Press, New York

Softcover reprint of the hardcover 1st edition 1982

A Division of Plenum Publishing Corporation

233 Spring Street, New York, N.Y. 10013

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

Contributors

Jaap J. Beintema Biochemisch Laboratorium, Nijenborgh 16, 9747 AG Groningen, The Netherlands

Michael Coates School of Australian Environmental Studies, Griffith University, Nathan, Queensland, Australia 4111

John Czelusniak Departments of Biology and Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48202

Margaret O. Dayhoff National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. 20007

Wilfried W. De Jong Department of Biochemistry, University of Nijmegen, Geert Grooteplein N21, 6525 EZ Nijmegen, The Netherlands

Howard Dene Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201

Morris Goodman Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201

David Hewett-Emmett Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109

Richard Holmquist Space Sciences Laboratory, University of California at Berkeley, Berkeley, California 94720

Lois T. Hunt National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. 20007

Thomas H. Jukes Space Sciences Laboratory, University of California at Berkeley, Berkeley, California 94720

Johannes A. Lenstra Biochemisch Laboratorium, Nijenborgh 16, 9747 AG Groningen, The Netherlands

Michael J. Novacek Department of Zoology, San Diego State University, San Diego, California 92182

Dennis Pearl Department of Statistics, University of California at Berkeley, Berkeley, California 94720

Alejo E. Romero-Herrera Department of Anatomy, Wayne State University School of Medicine, Detroit, Michigan 48201

Alan F. Scott The Johns Hopkins University School of Medicine, and Howard Hughes Medical Institute, Laboratory for Human Biochemical Genetics, Baltimore, Maryland 21205

Kirby D. Smith The Johns Hopkins University School of Medicine, and Howard Hughes Medical Institute, Laboratory for Human Biochemical Genetics, Baltimore, Maryland 21205

Simon Stone Department of Botany, University of Adelaide, Adelaide, South Australia 5001

Richard E. Tashian Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109

Patrick J. Venta Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109

Preface

George H. F. Nuttall pioneered the study of phylogeny through the genetically encoded sequence structures of proteins. His classic monograph, *Blood Immunity and Blood Relationship*, was published in 1904. The findings described in this monograph testified that immunologic comparisons of serum proteins could help reveal the phyletic relationships of primates and other animals. Although Nuttall had no way of knowing that a correspondence between the nucleotide sequences of genes and the amino acid sequences of proteins was the genetic basis for the immunologic specificities of animal sera, he clearly saw the implications of his findings. Thus he wrote in the introduction of his monograph, "The persistence of the chemical blood-relationship between the various groups of animals serves to carry us back into geological times, and I believe we have but begun the work along these lines, and that it will lead to valuable results in the study of various problems of evolution." Nuttall's prophecy is being fulfilled. Through the first two-thirds of the 20th century immunology led the way in the molecular analysis of the phyletic relationships of animal taxa above the species level. Amino acid sequencing of proteins began in earnest during the 1960s. It overtook immunology during the 1970s and provided more exact molecular data for investigating the history of life and the forces of chance and selection which drive evolution. The dramatic development of recombinant DNA and nucleotide sequencing technologies are now bringing us to a new watershed in comparative molecular studies. A far more detailed reconstruction of the history of genes and of the phylogeny of organisms containing the genes than hitherto attempted is becoming possible. Genealogical reconstructions carried out by the parsimony method on considerable amino acid sequence data, as well as on the initial body of globin gene nucleotide sequence data, are

already providing valuable results for systematic and evolutionary biology. These results are described in the present volume.

This volume is based on a symposium which explored how sequence data from proteins and DNA are used in studying phylogeny and evolution. This symposium was held at the Third International Congress of Systematic and Evolutionary Biology during the summer of 1980 in Vancouver, British Columbia. In addition to the chapters based on papers presented at this symposium, several further contributions are included, one to provide background information on placental mammalian evolution and two for wider coverage of new directions made possible by recombinant DNA and nucleotide sequencing technologies. The various orders of placental mammals in the Eutheria are presently better represented by amino acid sequence data than are any other major group of eukaryotes. Thus these data, which come from amino acid sequencing mainly of ribonucleases, lens α -crystallins, myoglobins and α - and β -hemoglobins, cytochromes *c*, fibrinopeptides A and B, and carbonic anhydrases, are being used to investigate eutherian systematics. In turn, evidence from the fossil record on times of divergence between eutherian lineages is testing the clock model of molecular evolution and provides data needed for investigating tempo and mode in molecular evolution. Our vision, of course, should not be limited to the Eutheria. Sequence data on certain proteins and genes can encompass the whole panorama of living species and thus provide the means for investigating the history of life from its earliest stages on earth to the present. That the study of phylogeny and evolution through protein and DNA sequences is very much in its infancy, despite a vigorous beginning, should readily be apparent to the readers of this volume, as it is to the authors. The best years lie ahead.

Morris Goodman
Wayne State University

Contents

Part I. Eutherian Phylogeny and Protein Evolution

Chapter 1

Information for Molecular Studies from Anatomical and Fossil Evidence on Higher Eutherian Phylogeny

Michael J. Novacek

1. Introduction	3
2. A Historical Sketch.....	5
3. Eutherian Monophyly	10
4. Higher Level Eutherian Phylogeny.....	12
4.1. Edentata, Pholidota	16
4.2. Carnivora, Creodonta	19
4.3. Rodentia, Lagomorpha, Macroscelidea	20
4.4. “Archontans” and Insectivora	21
4.5. The Ungulate Radiation: Perissodactyla, Artiodactyla, Cetacea, Proboscidea, Tubulidentata, Hyracoidea, Sirenia	25
5. Divergence Times for Eutherian Orders	27
6. Eutherian Phylogeny, Fossil History, and Evolutionary Rates	30
Appendix	34
References	35

Chapter 2

Evolution of Mammalian Pancreatic Ribonucleases

Jaap J. Beintema and Johannes A. Lenstra

1. Introduction	43
2. Trees.....	44
2.1. Most Parsimonious Tree	48
2.2. Biologic Tree.....	50
3. Evolutionary Rates	52
4. Three-Dimensional Structure of Ribonuclease.....	56
5. Enzymic Activity on Low-Molecular-Weight Nucleotide Substrates and on RNA	58
6. Enzymic Activity on Double-Stranded RNA	61
7. Interaction with Inhibitor Proteins	64
8. Glycosylation and the Function of Carbohydrate	64
References	71

Chapter 3

Eye Lens Proteins and Vertebrate Phylogeny

Wilfried W. De Jong

1. Introduction	75
2. Comparative Anatomy and Evolution of the Lens	75
3. Comparative Studies of Lens Proteins	78
3.1. The Crystallins.....	78
3.2. Intraspecies Variation in Lens Proteins.....	82
3.3 Crystallin Variation between Species	82
4. Structural Analysis of α -Crystallin	84
4.1. Isolation of α -Crystallin	84
4.2. Electrophoresis of the α -Crystallin Chains	87
4.3. Sequence Analysis of α A Chains.....	88
4.4. Phylogenetic Reconstructions	91
5. Phylogenetic Inferences from α -Crystallin A Sequences	93
5.1. Mammalian Phylogeny	93
5.2. Relationships among Vertebrate Classes and Subclasses.....	104
6. Molecular Aspects of α -Crystallin Evolution	106
6.1. Parallel and Back Substitutions.....	108
6.2. Changes in Charge Are Avoided	108
6.3. "Covarions"	109
6.4. Variable Rates of Change and a Directional Trend in Substitutions	109
6.5. Unequal Distribution of Substitutions over the α A Chain....	110
References	112

Chapter 4**Amino Acid Sequence Evidence on the Phylogeny of Primates and Other Eutherians****Morris Goodman, Alejo E. Romero-Herrera, Howard Dene, John Czelusniak, and Richard E. Tashian**

1. Introduction	115
2. Genealogic Evidence from Amino Acid Sequences	116
2.1. Tree Construction Strategy	116
2.2. Sequences Analyzed	122
2.3. Gnathostome Myoglobin Genealogy	130
2.4. Genealogic Evidence from Other Gene Phylogenies	136
2.5. Species Phylogeny from Combined Sequence Data	145
3. Utilizing the Clock Model of Protein Evolution	148
3.1. Rationale and Procedure	148
3.2. Results Obtained by the Clock Model	152
4. The Tempo and Mode of Protein Evolution	161
4.1. Pattern of Rate Variations	161
4.2. The Central Role of Natural Selection	167
5. Conclusions	183
Epilogue: New Cladistic Findings on Globin Phylogeny	185
References	187

Chapter 5**Evolution of Chromosomal Proteins****Lois T. Hunt and Margaret O. Dayhoff**

1. Introduction	193
2. Nucleosome Structure	194
2.1. The Chromatin Fiber	194
2.2. The Nucleosome and Nucleosome Core	195
3. Computer Methods	199
3.1. Detection of Protein Relationships	199
3.2. Protein Alignments and Evolutionary Trees	200
3.3. Protein Families and Superfamilies	202
3.4. Mutation Acceptance Rates	203
4. Chromosomal Proteins	203
4.1. Prokaryote DNA-Binding Proteins	203
4.2. Viral Nucleic Acid-Binding Proteins	205

4.3. Protamines	206
4.4. Nonhistone Chromosomal Proteins	207
4.5. Histones H1 and H5.....	212
4.6. Histone H2A and Nuclear Protein A24	212
4.7. Histone H2B	217
4.8. Histone H3.....	219
4.9. Histone H4.....	221
4.10. Mutation Acceptance Rates	224
4.11. The Core Histone Superfamily	225
References	230

Part II. Modeling the Process of Sequence Divergence

Chapter 6

Simulation of the Evolution of Macromolecular Sequences by Random Fixation of Allowed Codons

Michael Coates and Simon Stone

1. Introduction	243
2. The Simulation Model—RFAC	247
2.1. The Input Data	247
2.2. The Starting DNA Sequence	247
2.3. Random Point Mutation.....	248
2.4. Allowed Amino Acids.....	248
2.5. Fixation	251
2.6. Phylogenies	251
3. Comparison of RFAC with Other Models	254
4. Methods of Comparison of Macromolecular Sequence Divergence in Real and Simulated Evolution.....	256
5. Results.....	257
5.1. Effects of Rate of Fixation and Amount of Allowed Variation on Macromolecular Sequence Divergence	257
5.2. Fitting of Simulated to Real Results of Evolution.....	258
5.3. Effect of Amount of Allowed Variation on “Efficiency” of Evolution	269
6. Conclusions and Discussion	273
References	277

Chapter 7**Nonuniform Molecular Divergence: The Quantitative Evolutionary Analysis of Genes and Messenger RNAs under Selective Structural Constraints****Richard Holmquist, Dennis Pearl, and Thomas H. Jukes**

1. Introduction.....	281
2. Constrained Stochastic Theory.....	282
3. Magnitude of Selective Constraints	284
3.1. Base Composition	284
3.2. Base Replacement Probabilities	285
3.3. Density of Fixed Mutations among Variable Codons.....	286
3.4. Distribution of Fixed Mutations within Codons.....	288
4. Estimation of Parameters	291
4.1. The Observational Principle Underlying Parameter Estimation in Gene Sequences	291
4.2. Evolutionary Measures	294
5. A Comparison of Evolutionary Estimates Made from Protein and Nucleic Acid Sequence Data and from Nonrandom REH Theory	296
5.1. A Closely Related Divergence: Mouse and Rabbit β -Hemoglobin	296
5.2. A Distantly Related Divergence: Rabbit α - and β -Hemoglobin	298
6. Comparison of Calculation with Experiment.....	299
7. Generality of Results	302
8. Extension to Noncoding Regions	305
9. Importance of Accurate Genetic Distance Estimates to Systematics	306
10. Discussion	309
11. Conclusions.....	312
References	312

Part III. Prospects for Investigating Evolution
through Genomic DNA

Chapter 8**Genomic DNA: New Approaches to Evolutionary Problems****Alan F. Scott and Kirby D. Smith**

1. Introduction.....	319
2. Overview of Methods	320

3. Comparative Studies of Genomic DNA	327
3.1. Y-Chromosome DNA	327
3.2. Other Middle-Order Reiterated DNAs	329
3.3. Studies of the Ribosomal Genes	331
3.4. Studies of Single-Copy DNA	336
3.5. DNA Polymorphisms	344
3.6. Other DNAs	346
4. Conclusions	349
References	350

Chapter 9

Features of Gene Structure, Organization, and Expression
That Are Providing Unique Insights into Molecular Evolution
and Systematics

David Hewett-Emmett, Patrick J. Venta, and Richard E. Tashian

1. Introduction	357
2. Recombinant DNA Technology	358
2.1. Cloning	358
2.2. DNA and RNA Sequencing	362
3. Expression and Organization of Eukaryotic Genes	364
3.1. DNA Transcription and Processing of RNA	364
3.2. The Intervening Sequences of Genes	367
3.3. The Role of Intervening Sequences	369
3.4. Gene Duplication	373
3.5. Pseudogenes	377
4. Use of Nucleotide Sequences in Evolutionary Tree-Building	382
4.1. Why Nucleotide Sequences?	382
4.2. Tree-Building Strategy	384
4.3. A Hemoglobin Nucleotide Tree	384
5. Concluding Remarks	397
References	398

Index	407
-------------	-----